DOCUMENT RESUME

ED 162 856 SE 025 302

AUTHOR Staver, John R.; Gabel, Dorothy L.
TITLE The Development and Construct Validation of a Group Administered Test of Piaget's Formal Thought.
PUB DATE 78
NOTE 24p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Cognitive Development; *Educational Research; *High School Students; Learning Theories; *Measurement Techniques; Reliability; *Science Education; Science Teachers; Validity
IDENTIFIERS *Piaget (Jean); Research Reports

ABSTRACT

This study investigates the reliability and construct validity of a group administered test of Piaget's formal operations stage. A related problem involving a learning effect associated with Piaget's clinical methods is also investigated. The Piagetian Logical Operations Test (PLOT), a group-administered instrument, was developed and field-tested. Eighty-four students in grades 10-12 participated in the field test. Subjects were randomly selected and assigned membership in one of two equal size groups having the same number of males and females from each grade. Data were obtained by clinical interview, PLOT, and intelligence test records. Subjects in group one received five clinical interviews followed by PLOT, while subjects in group two were administered the instruments in reverse order. Examination of the findings led to the conclusions: (1) the construct validity of the group test was partially established; (2) a learning effect was present in the PLOT total scores which was attributable to the previously administered clinical interviews, but no such effect was present, in general, in the clinical interview scores that were attributable to the administration of PLOT. (Author/HM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED162856

The Development and Construct Validation of a Group Administered Test of Piaget's Formal Thought

John R. Staver

De Paul University

Dorothy L. Gabel

Indiana University

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John R. Staver

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Running head: Construct Validation

SE 025 302

Abstract

The problem of this investigation was to answer the question: Can a group-administered test of Piaget's formal operational stage be developed and construct validated? A related problem involving a learning effect associated with Piaget's clinical methods was also investigated. The Piagetian Logical Operations Test (PLOT), a group-administered instrument, was developed and field-tested to answer the question of this investigation.

Eighty-four students in grades 10 - 12 of a south central Indiana consolidated school corporation participated in the field test. Subjects were randomly selected and assigned membership in one of two equal size groups having the same number of males and females from each grade. Data, which was obtained by clinical interview, PLOT, and intelligence test records, was analyzed using a 2 x 3 x 2 factorial design. Subjects in group one received five clinical interviews followed by PLOT while subjects in group two were administered the instruments in reverse order. A Campbell and Fiske multitrait-multimethod matrix consisting of three methods and four traits, factor analysis, and three-way ANOVA were employed to statistically examine the date obtained.

Analysis of data revealed several findings: (1) The internal consistency reliability (alpha) of PLOT was .85. Reliability of individual scales was also reported. (2) PLOT was significantly and substantially correlated with Piaget's clinical method. (3) PLOT total scores and intelligence test scores did not show high factor loadings on the same factor. PLOT total scores and clinical interview total scores did not exhibit high factor loadings on the same factor. (4) Subjects who were previously administered clinical interviews scored significantly higher on PLOT than subjects who did not receive interviews prior to PLOT, but subjects who receive PLOT previous to the clinical interviews did not score significantly higher on the total clinical interview score than subjects who did not take PLOT prior to the clinical interviews.

Examination of the findings led to two conclusions: (1) The construct validity of the group test was partially established. (2) A learning effect was present in the PLOT total scores which was attributable to the previously administered clinical interviews, but no such effect was present, in general, in the clinical interview scores that was attributable to the previous administration of PLOT.

Introduction

Piaget employed the clinical interview because that method provided the most useful framework for his research on the development of cognitive thought within the individual. However, educators who wish to study cognitive development and its implications for science teaching across individuals by the clinical method encounter two major drawbacks. One is the amount of time consumed, and the second is the inherent methodological nonstandardizations associated with the clinical method.

Several workers (Burney, 1974; Lawson, 1978; Longeot, 1963, 1964; Raven, 1973; Renner, 1977; Shayer and Wharry, 1974; and Tisher, 1971) attempted the construction of a group administered measure of Piagetian cognitive development. One goal of these assessments was the modification of science teaching strategies for better consistency with the intellectual development of children. However, each effort only partially meets three criteria which seem prerequisite for a valid efficient test: (1) logical equivalence of written test items and the mental logic of specific Piagetian tasks; (2) evaluation of the reliability and construct validity of the group measure; (3) assessment in an efficient objective format of specific reasons offered by children in support of cognitive decisions. The goal of this study is to describe the development and construct validation of a test which fulfills the aforementioned criteria.

The Piagetian Logical Operations Test (PLOT)

In this section several characteristics of the Piagetian Logical Operations Test (PLOT) are delineated, including format, items, scales, and scoring procedures. PLOT is an objective multiple-choice test with four alternatives per question and four individual scales: (1) conservation of volume by liquid displacement, (2) separation and control of variables, (3) combinatorial analysis, and (4) proportional thought. The conservation scale represents a trait of late concrete thought proposed by Karplus and Lavatelli (1969). The three remaining scales each represent a trait of formal thought proposed by Piaget (Indelder and Piaget, 1958). Each scale consists of three item types, content questions that assess the subject's comprehension

of a task, decision questions which require a cognitive decision by the student, and reason questions which identify reasons for cognitive decisions. At least one reason question is designed to specifically rate subject reasoning patterns on each decision question. All PLOT questions are similar to questions asked in clinical interviews, the principal difference being the format. Thus, the logic necessary to answer the questions may be assumed identical to the logic required to solve the corresponding clinical tasks.

At least one cognitive task appraising each trait of Piagetian thought is presented via video-tape. The same tasks were also given by clinical interview, and they are described in that section of the report. Employment of video-tape demonstrations of Piagetian tasks permit the administration of PLOT to classroom size groups (30 students) and control of variation in administration procedures. Subjects observe the task and answer questions in the appropriate section of the test booklet. A PLOT total score and individual PLOT scale scores are available, and each score is calculated by summing the number of correct answers in the appropriate scale or the entire test.

Validation Procedures

The procedures employed to evaluate the reliability and construct validity of PLOT are described in this segment. Included are the aspects of construct validity, instruments, statistical procedures, and special problems associated with Piagetian measurement.

Construct validation is typically a two dimensional process. One aspect, convergence, is concerned with sustainment by independent measurement, and the other dimension, descriminance, is focused on the independence of tests not constructed to measure the same traits (Nunnally, 1967). To examine both aspects of construct validity, multiple traits and multiple methods must be employed (Campbell and Fiske, 1959).

The methods utilized in this study were PLOT, the Piagetian clinical interview, the Lorge-Thorndike Intelligence Test, (Form 1, Levels C,D,E) and the Cognitive Abilities Test, (Form 1, Level G). Traits measured were conservation of volume by liquid displacement, separation and control of variables, combinatorial analysis, proportional thought, verbal, nonverbal, and quantitative abilities.

The Lorge-Thorndike Intelligence Test and the Cognitive Abilities Test are measures of a general mental ability, the former having verbal and nonverbal scales, and the latter having verbal, nonverbal, and quantitative scales. A score for each scale was used as well as a total score for each mental ability measure, the sum of scales scores for the C.A.T. and the mean of the scales for the L.T.I.T.

Five Piagetian tasks were selected for administration to subjects by clinical interview: (1) Volume of Metal Cylinders by Liquid Displacement (Karplus and Lavatelli, 1969); (2) Flexibility of Bending Rods (Inhelder and Piaget, 1958); (3) Colored and Colorless Chemicals (Inhelder and Piaget, 1958); (4) Mr. Tall-Mr. Short-Measurement with Paper Clips (Karplus and Lavatelli, 1969); (5) Equilibrium in the Balance (Inhelder and Piaget, 1958). Tasks 1,2, and 3 assess conservation of volume, separation and control of variables, and combinatorial analysis, respectively whereas tests 4 and 5 measure direct and inverse aspects of proportional thoughts, respectively.

Two evaluations of each clinical interview were made. First, a categorical (yes/no) decision concerning the presence of a mental schema was made. Second, a series of behavior statements representing possible behaviors of subjects during interviews were marked (yes=1/no=0) and totaled. Lists of behavior-oriented statements, called behavior observation sheets, were previously found to be reliable and valid by Staver (1977) in measurement of Piagetian schema by clinical interviews. Evaluations of the inter-rater reliability and concurrent validity of behavior observation sheets employed in this research are discussed elsewhere (Staver 1978).

All clinical interview evaluations were done by a three-judge panel of advanced science education graduate students and post-doctoral research associates. Training of evaluators included discussion of involved schema, clarification of behavior statements, and practice in the use of the behavior observation sheets.

Employment of several measurement methods and assessment of several traits of cognitive thought, although necessary for evaluation of construct validity, can become unwieldy. A convenient way to simplify the evaluation is to construct a multi-trait-multimethod matrix of the correlations. The Campbell and Fiske (1959) model used in this study is a presentation of all correlations among several traits and methods in matrix form for such evaluation.

To further evaluate the construct validity of PLOT, the scores of all instruments were subject to a factor analysis. Results of this procedure could provide additional evidence for convergence and discriminance. The SPSS-Factor Program (Nie, et al., 1975), employing the principal components method with iterations to achieve orthogonal factors and varimax rotation to simple structure of all factors having eigenvalues greater than 1.0, was used.

Two validation problems remain to be delineated, the selection of a sample and the evaluation of a learning effect associated with the clinical method. A critical aspect of the validation procedure was the selection of a subject sample from a population containing substantial numbers of concrete, transitional and formal thinkers. Based upon Chiappetta's (1976) review of studies concerning the developmental levels of secondary and college students, the conclusion was made that a random sample of senior high school subjects would provide the best mixture. Therefore twenty-one males and twenty-one females were randomly selected from each grade of a large 10-12 grade high school in a south central Indiana consolidated school corporation. This selection procedure yielded a sample of 126 subjects which contained equal numbers of males and females within each grade.

The final validation problem was a learning effect associated with the clinical method. Subjects often show more advanced reasoning patterns in the second interview when a clinical task is administered twice within a brief time period. Such learning effects could act to decrease correlations among Piagetian variables and give spurriously low estimates of convergent validity. To evaluate subject learning effects, seven males and seven females within each grade were randomly assigned membership in the cells of a 2x3x2 factorial design (Kirk, 1968) and a replacement group involving two groups, three grades, and two sexes. No pretest was employed because of the reactivity of Piagetian measures. Further, Campbell and Stanley (1963) maintain that the most adequate assurance concerning the absence of initial bias between groups is randomization. Treatment was considered to be the administration of the five clinical tasks in order 1, 2, 3, 4, 5, and PLOT was considered to be the posttest. The 42 students comprising group 1 were given treatment before posttest evaluation whereas an equal number of subjects in group 2 were administered the posttest followed by treatment. Thus, each group acts as a control for its counterpart. The 42 subjects in group 3 formed a replacement pool. Children failing to participate in the first activity of their assigned group were replaced by a randomly chosen subject whose grade and sex matched that of the lost subject. No student who failed to continue after participating in the initial activity was replaced.

Findings, Conclusions, and Discussion

Validity of PLOT

To evaluate the reliability and construct validity of PLOT, information derived from the corralational matrix, factor analysis, the learning effect, and the efficiency of PLOT is set forth in succeeding parts of this section.

The correlations among the three methods and four traits are assembled into a Campbell and Fiske matrix and presented in Table 1. A detailed inspection of the

(Insert Table 1 about here)

matrix is necessary to determine the findings. The internal consistency reliability (alpha) value for each instrument scale is shown as the value in parenthesis. For example alpha=.85 for the PLOT conservation scale. Alpha for PLOT total score, not shown in Table 1, is also .85. According to criteria set forth by Davis (1964) for individual differences measurement, the reliabilities of PLOT scales 1 and 4, and the total score are acceptable whereas alpha for PLOT scales 2 and 3 are insufficient.

Four criteria are examined in Table 1 to determine the validity of PLOT. First, correlations of the same trait measured by different methods should be significant and substantial. These correlations form three diagonals called validity diagonals and the entries are all underscored. Seven of the twelve validity diagonal values are significant and substantial, thereby indicating convergence among the methods for those traits. Second, measures of the same trait should exhibit higher positive correlations between each other than with measures of different traits employing different methods. This means that a validity diagonal entry in Table 1 should be greater than values in its row and column of the adjacent heterotrait-heteromethod triangles (enclosed by broken lines). Inspection of Table 1 for the seven significant validity diagonal cases shows the second criterion fulfilled in only two cases, B_1B_2 and D_1D_2. Third, measures of the same trait should show higher positive correlations between each other than with measures of different traits using the same method. With respect to Table 1, the validity diagonal value for a variable should be higher than its values in the heterotrait-monomethod triangles (enclosed by solid lines). Examination shows that only one significant validity case, D_1D_2, meets this criteria. Fourth, measures of different traits should exhibit an identical pattern of intercorrelations among each other across heterotrait-monomethod and heterotrait-heteromethod triangles. Such a pattern in Table 1 would be a single trend in the magnitudes of correlations for all triangles. In fact, no single pattern or trend is detected. The last three criteria are focused on the discriminant aspect of con-

struct validity, and analysis of Table 1 shows little evidence of discriminant validity for the Piagetian and general intelligence measures.

Factor analysis represents an additional method for the evaluation of the construct validity of PLOT, and the results of PLOT, clinical interviews, and the mental ability tests are shown in Table 2. Convergent validity between variables is exhibited by high loadings for variables on the same factor whereas discriminant validity among variables is supported by high loadings coupled with modest loadings on the same factor (modest-high couple). In this study a high loading is $\geq .60$, a medium loading is $\geq .40$ and $\leq .59$, and a low loading is $\leq .39$. Inspection of Table 2 reveals that high loading on the same factor are not observed for PLOT total score and all PLOT scale scores with the corresponding total clinical interview score and clinical task scores. Therefore, little evidence for convergence between the two Piagetian methods is present. Modest-high factor loading couples for PLOT total scores, and PLOT scales, 1, 3, and 4 with intelligence test scores are observed whereas only half the modest-high couple is seen for PLOT scale 2 with intellegence measures; PLOT scale 2 exhibits a medium factor loading on factor 1 which exhibits high loadings for intelligence measures. Thus, substantial evidence for discriminant validity between Piagetian and general intelligence measures is found, but little evidence for convergence of the two Piagetian measures is observed.

The correlational and factor analytical findings present an enigmatic situation which requires discussion. The correlational analysis provides evidence only for convergence between Piagetian measures whereas the factor analysis provides rather clear evidence for discriminance, but little support of convergence is found. The correlation between the PLOT total score and the total clinical interview score, .59, is comparable with higher validity diagonal values in Table 1 and further supports convergence. However, the lowest correlations in the heterotrait-heteromethod triangles suggest that the measurement methods in this research are not entirely inde-

pendent. PLOT and the clinical interview method share common materials, tasks, and questions. Principal differences are demonstration versus manipulation of materials, forced multiple-choice versus open-ended question-answer format, and written versus oral response. PLOT and the mental ability tests share a common question-answer format and the necessity of reading for comprehension. Therefore, it is probable that all three methods are related. Additionally, the traits themselves may form a unified system of thought and are not completely independent. Campbell and Fiske (1959) maintain that some evaluation of validity can be made in this situation, and accordingly, some convergence is indicated for PLOT scales 2 and 4 and the PLOT total scores.

With respect to the factor analytical procedures, little or no evidence of convergent validity is found by observation of high-high loading couples on the same factor for PLOT and clinical interview variables. Further, substantial results indicating discriminant validity are present in the loading patterns of PLOT and mental ability variables. The loading patterns permit both the identification of rotated factors and the establishment of discriminant validity for PLOT by this method.

In the factor solution presented in Table 2 only mental ability variables exhibit high loadings on factor one. Remaining variables load modestly with one notable exception, PLOT part 2, on this factor; it shows a loading on factor one of .53, medium. Factor one is clearly identifiable as a factor associated with general intelligence. PLOT part 2 loads substantially on this factor because the ability to separate and control variables seems to be associated with general mental ability. This factor also accounts for 76.4% of the total variance.

Factor two, which accounts for 13.3% of the total variance, is somewhat more difficult to identify. Inspection of factor two reveals high loading for five of the ten clinical interview variables whereas one of the remaining five variables shows a medium loading and the other four exhibit modest loadings on factor two. All intelligence variables load modestly on this factor as well as do all PLOT variables.

Analysis of factor three, accounting for 10.3% of the total variance, further aids in the identification of factors two and three. All PLOT variables exhibit high loadings on factor three except scale 2 which loads .53, medium. Further, all mental ability and clinical interview variables are observed to load modestly on factor three. What seems to have occurred in the rotation to simple structure is a variable separation on orthogonal factors by method. Factor one, as previously identified, is associated with general mental ability. Factor two, although less clearly so, seems related to Piagetian cognitive development assessed through clinical interviews whereas factor three is revealed to be connected with PLOT as a measurement method of Piagetian cognitive development. Although the factor solution gives ample evidence of discriminant validity, it also yields little suggestion of convergence for PLOT and the clinical method. Therefore, it is concluded that convergent and discriminant validity are partially established.

Learning, Sex, and Grade Effects

Three-way analyses of variance were performed on the PLOT and clinical interview scores, and the findings, conclusions, and discussions of the learning phenomenon, plus grade and sex effects are set forth in this section.

Significant differences in favor of the group which was previously administered the series of clinical interviews exists in the PLOT total, scale 1, and scale 3 mean scores compared to the group which did not receive clinical interviews prior to PLOT administration ($F=12.06$, 15.90, 6.53, respectively; $p<.05$, $df=1,55$). Group mean differences for PLOT scales 2 and 4, although in favor of the group receiving prior clinical interviews, were not significant. Gradually increasing mean scores for PLOT and its individual scales were detected with increasing grade level, but mean differences were not significant. Also, no significant differences with respect to sex were revealed for PLOT and its scales, and no significant two and three-way interactions were present.

Group mean differences on the clinical interview variables, although favoring the group receiving PLOT prior to the interviews were generally not significant. Two exceptions were the significant differences on clinical tasks 1 and 2 ($F=12.84$, 6.96, respectively; $p<.05$, $df=1,60$; 1,58, respectively). An increase in grade level is generally accompanied by an increase in the mean for the total clinical score and all task scores, but only the mean difference for task 3 is significant ($F=3.94$; $p<.05$, $df=1,56$). The mean differences for sexes were not significant, and no significant two or three-way interactions were detected.

Consideration of the findings concerning learning effects leads to several conclusions. First, a learning effect attributable to the prior administration of clinical interviews is present in the PLOT total score and PLOT scales 1 and 3. Second, the clinical interviews, when considered as treatment, have a similar result across the main effects of group, grade, and sex taken in pairs or in triplet. Third, a learning effect attributable to prior PLOT administration is present only in the scores of clinical tasks 1 and 2. It is not present in any remaining clinical variable including the total score. Fourth, PLOT, when viewed as treatment, has a similar effect across the main effects of group, grade, and sex taken in pairs or in triplet.

The general presence of a learning effect in PLOT scores attributable to the prior administration of clinical interviews, and the general absence of such an effect in the clinical scores due to prior PLOT administration presents another enigmatic situation. The two instruments are designed to measure the same traits, and they have common materials and similar questions. A plausible explanation arises from the theory itself, and Piaget's thoughts on the self-regulation mechanism. Prior to the onset of formal operations, and still valuable in formal thought, is the active manipulation of the environment by the child. A fundamental difference between PLOT and the clinical tasks is that during interviews subjects actively manipulate the materials whereas such objects are only observed on video-tape on PLOT. This manipulation-observation difference seems important in accounting for the

general presence of learning associated with the clinical method and its general absence in PLOT. The isolated cases of learning in clinical tasks 1 and 2 attributable to prior PLOT administration are most probably explained by shared trait and method variance.

Ex Post Facto Analysis

Data analysis, including item analysis of PLOT, indicated three areas for ex post facto examination of the data concerning the reliability and construct validity of PLOT. The areas are correction for attenuation in correlations, deletion of PLOT content questions, and evaluation of PLOT decision and reason scales.

The correction for attenuation procedure (Guilford and Fruchter, 1978) was applied to each entry in Table 1 to determine the extent of the detrimental effect exerted on the validity of PLOT by the low reliability coefficients for PLOT scales 2 and 3. Although the unattenuated correlations were higher, especially entries in the validity diagonals, no new information about the construct validity of PLOT was yielded by the analysis method described earlier. Therefore, it was concluded that while the low reliabilities of the two PLOT scales are detrimental, they are not the primary problem in establishing the construct validity of PLOT.

Three kinds of items, content, decision, and reason, compose PLOT. Item analysis revealed that students obtained a mean of 11.97 and a standard deviation of 1.05 on the thirteen content questions whereas their performance was much more diverse on the decision and reason questions. The deletion of content questions from the PLOT total score represents an approximate linear transformation. Such transformations have no effect on correlation among variables; (Hopkins and Glass, 1978) thus removal of content questions has little effect on construct validity. Deletion of PLOT content questions did yield a slight positive trend in reliability. Such questions could be deleted from the entire test, but some discussion is justified concerning the role of content items in the measurement process.

One line of thought is to remove such items from the test altogether because they serve no other function than to increase the score. A second direction for consideration is to remove the content items only from the score because the function of such items is to focus the subjects' attention on the most important aspects of the problems to be solved. This point is crucial because the subjects only view a demonstration of the problems; materials are not handled. Therefore, the presence of such items may be critical to the subjects' comprehension of the problems, and it thereby influences answers to decision and reason questions. The fact that most subjects receive a near perfect score merely indicates that the goal for which the questions are designed is being achieved. Thus, a student's score on the content items provides little indication of current developmental level. That information is yielded through answers to decision and reason questions in each scale.

A third direction of ex post facto analysis seemed justified. The concepts 'decision' and 'reason' were considered as traits measured by the three aforementioned methods in a new Campbell and Fiske matrix and the correlations appear in Table 3.

(Insert Table 3 about here)

Analysis of Table 3 by methods outlined earlier revealed substantial evidence for convergence among all the methods (all validity diagonal entries are significant and most are substantial), but little information concerning discriminance. Thus, no new findings were uncovered, and the previous discussion of the matrices holds. In summary, ex post facto analysis yielded no information which conflicted with earlier results.

Efficiency and Practicality of PLOT

One requirement cited earlier for a useful Piagetian test was the development of an efficient practical measure. An important characteristic of PLOT as an untimed test is that each group proceeds through the sequence of video tape demonstrations and written questions at the pace of the slowest student. Data concerning time

required for administration of PLOT to subject show that PLOT was administered 23 times with a mean administration time of 46.1 minutes and a range of 38 - 56 minutes. These data are indicative of the fact that PLOT can be administered to individuals or small groups of students within a 55-minute period.

Implications for Teachers

The development and construct validation of PLOT, a group measure for assessing four Piagetian schema associated with formal thought was reported in this paper. PLOT was developed for use by science teachers and researchers in science education interested in the assessment of developmental reasoning capabilities of students. One goal of science teaching is to match instruction and curriculum materials with the developmental level of the learner. Learning difficulties of students in middle and secondary school science have often been attributed to an inability to grasp concepts in science. A more refined line of thought suggests that some students are not yet using reasoning patterns required to comprehend certain science concepts. Furthermore, many concepts in science may be taught in a manner consistent with either formal or concrete thought. However, a prerequisite to the matching process is a reliable, valid, efficient and practical measurement device. Although further test development of PLOT is appropriate, the preponderance of evidence suggests that PLOT is a reliable, valid, efficient, and practical measurement tool, and may thus be employed by teachers and researchers for the aforementioned purposes.

References

Burney, G. M. The construction and validation of an objective formal reasoning instrument. (Doctoral dissertation, University of Northern Colorado, 1974). (University Microfilm No. 75-5403).

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrate-multimethod matrix. Psychological Bulletin, 1959, 56(2), 81-105.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1963.

Chiappetta, E. L. A review of Piagetian studies relevant to science instruction at the secondary and college level. Science Education, 1976, 60(2), 253-261.

Davis, F. B. Educational measurements and their interpretations. Belmont, California: Wadsworth Publishing Co., 1964.

Guilford, J. P., & Fruchter, B. Fundamental statistics in psychology and education (6th Ed.). New York: McGraw-Hill, 1954.

Hopkins, K. D., & Glass, G. V. Basic statistics for the social sciences. Englewood Cliffs, New Jersey: Prentice-Hall, 1978.

Inhelder, B., & Piaget, J. The growth of logical thinking from childhood to adolescence; an essay on the construction of formal operational structures. New York: Basic Books, 1958.

Karplus, R., & Lavatelli, C. The developmental theory of Piaget: Formal thought. San Francisco: John Davidson Film Producers, 1969. (Film)

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.

Lawson, A. E. The development and validation of a classroom test of formal reasoning. Journal of Research in Science Teaching, 1978, 15(1), 11-24.

Longeot, F. [An Essay of the application of genetic psychology to differential psychology.] B.I.N.O.P. (Bulletin De L'Institute D'Etude Du Travail Et D'Orientation Professionalle), 1962, 18 153-162.

Longeot, F. [Statistical analysis of three collective genetic tests.] B.I.N.O.P. (Bulletin De L'Institute D'Etude Du Travail Et D'Orientation Professionalle), 1964, 20, 219-232.

Nie, N. H., Hull, C. H., Jenkins, J. G., Speinbrenner, K., & Bent, D. H. Statistical package for the social sciences (2nd Ed.). New York: McGraw-Hill, 1975.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.

Raven, R. J. The development of a test of Piaget's logical operations. Science Education, 1973, 57, 377-385.

Renner, J. W. Evaluating intellectual development using written responses to selected science problems. A report to the National Science Foundation on Grant No. EPP75-19596, Analysis of Cognitive Processes, University of Oklahoma, Norman, 1977.

Shayer, M., Wharry, D. Piaget in the classroom part 1: Testing a whole class at the same time. School Science Review, March, 1974, 55(192), 447-458.

Staver, J. R. A testing of the waters of formal thought development. Teacher Education Forum, 1977, 5(8), 1-16.

Staver, J. R. The development and construct validation of a group administered test of Piaget's formal thought. (Doctoral dissertation, Indiana University, 1978).

Tisher, R. P. A Piagetian questionnaire applied to pupils in a secondary school. Child Development, 1971, 42, 1633-1636.

TABLE 1

CORRELATIONS AMONG FOUR TRAITS OF FORMAL THOUGHT MEASURED BY THE PIAGETIAN LOGICAL OPERATIONS TEST, CLINICAL INTERVIEWS AND THE COGNITIVE ABILITIES TEST

		PLOT				CLINICAL INTERVIEW				CAT			
		A_1	B_1	C_1	D_1	A_2	B_2	C_2	D_2	A_3	B_3	C_3	D_3
PLOT (N = 66)													
Cons. of Vol. by Liq. Dipl.	A_1	(.85)											
Sep. & Control of Variables	B_1	.47	(.56)										
Combinatorial Analysis	C_1	.32	.57	(.46)									
Proportional Thought	D_1	.42	.39	.52	(.78)								
CLINICAL INTERVIEW (N = 72)													
Cons. of Vol. by Liq. Dipl.	A_2	.07	.20	.21	.37	(.11)							
Sep. & Control of Variables	B_2	.19	.50*	.34	.29	.50	(.85)						
Combinatorial Analysis	C_2	.10	.41	.18	.02	.07	.27	(.79)					
Proportional Thought	D_2	.19	.47	.35	.58*	.30	.41	-.05	(.73)				
COGNITIVE ABILITIES TEST (N = 70)													
Verbal	A_3	.11	.62	.44	.22	.20	.46	.22	.49	(.94)			
Nonverbal	B_3	.07	.51*	.46	.42	.31	.44*	.10	.69	.70	(.90)		
Quantitative	C_3	.19	.58	.56*	.41	.28	.46	.05	.48	.74	.67	(.90)	
Total ($A_3 + B_3 + C_3$)	D_3	.14	.64	.55	.39*	.30	.51	.14	.62*	.91	.88	.90	(.87)

Note: Validity diagonals are the three sets of underlined values. Reliability diagonals are the three sets of values in parentheses. Each heterotrait-monomethod triangle is enclosed by a solid line. Each heterotrait-heteromethod triangle is enclosed by a broken line.

*significant, $p < .05$, one-tailed test.

TABLE 2

ROTATED FIVE FACTOR PATTERN OF PLOT, CLINICAL INTERVIEW AND MENTAL ABILITY TEST SCORES

Variable	Factors 1	2	3	4	5	Communality	N
Non Verbal IQ	.84	.14	.00	.17	.10	.77	68
Verbal IQ	.78	.06	.25	.18	.21	.74	68
Total IQ	.92	.10	.14	.19	.17	.93	68
CAT-Verbal	.87	.13	.09	.17	.15	.84	70
CAT-Quantitative	.71	.22	.15	-.03	.43	.75	70
CAT-Non verbal	.80	.23	.29	-.04	.08	.79	70
CAT-Total	.91	.22	.19	.03	.25	.97	70
Categorical Decision # 1	.05	.83	.23	.05	.07	.76	70
Categorical Decision # 2	.39	.58	.17	.31	.15	.63	68
Categorical Decision # 3	.12	.11	.05	.85	.13	.78	66
Categorical Decision # 4-5	.31	.14	.08	.07	.84	.83	70
Total Categorical Decision	.37	.60	.18	.49	.46	.99	66
Task 1	.10	.84	.07	-.05	.10	.73	66
Task 2	.35	.67	.11	.33	.13	.71	72
Task 3	.09	.09	.10	.81	-.11	.69	70
Task 4-5	.35	.22	.23	.01	.82	.91	68
Total Clinical Interview	.38	.62	.20	.51	.39	.99	66
PLOT-Part 1	.03	.01	.70	.14	.00	.51	66
PLOT-Part 2	.53	.15	.53	.38	.13	.74	66
PLOT-Part 3	.39	.21	.61	-.14	.05	.58	66
PLOT-Part 4	.07	.33	.66	-.07	.38	.69	66
PLOT-Total	.31	.22	.92	.13	.18	1.00	66
Eigen value	10.67	2.31	1.79	1.52	1.08		
Percent of variance accounted for	61.4	13.3	10.3	8.7	6.2		

NOTE: Principal components analysis with iterations, varimax factor rotation, and pairwise deletion of missing data was employed. The number of rotated factors was limited to five.

TABLE 3

CORRELATIONS AMONG THE DECISION AND REASON CHARACTERISTICS OF FORMAL THOUGHT MEASURED BY THE PIAGETIAN LOGICAL OPERATIONS TEST, CLINICAL INTERVIEWS, AND THE LORGE-THORNDIKE INTELLIGENCE TEST

		PLOT		Clinical Interview		Lorge-Thorndike Intelligence Test	
		A_1	B_1	A_2	B_2	A_3	B_3
PLOT	Decision A_1	(.61)					
	Reason B_1	.78	(.78)				
Clinical Interview	Decision A_2	.56*	.39	(.68)			
	Reason B_2	.58	.47*	.63	(.74)		
Lorge-Thorndike Intelligence Test	Verbal A_3	.56*	.45	.57*	.48	(.90)	
	Nonverbal B_3	.37	.30*	.51	.45*	.80	(.91)

*Note: Significant, $p < .05$, one-tailed test